



HAL
open science

Comparison of strategies to efficiently combine repeated urine samples in biomarker-based studies

Claire Philippat, Antonia M Calafat

► **To cite this version:**

Claire Philippat, Antonia M Calafat. Comparison of strategies to efficiently combine repeated urine samples in biomarker-based studies. *Environmental Research*, 2021, 192, pp.110275. 10.1016/j.envres.2020.110275 . inserm-03269506

HAL Id: inserm-03269506

<https://www.hal.inserm.fr/inserm-03269506>

Submitted on 24 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Comparison of strategies to efficiently combine repeated urine samples in biomarker-**
2 **based studies**

3

4 Claire Philippiat¹, Antonia M. Calafat²

5 *¹University Grenoble Alpes, Inserm, CNRS, Team of Environmental Epidemiology applied to*
6 *Reproduction and Respiratory Health, Institute for Advanced Biosciences (IAB), Grenoble,*
7 *France*

8 *²Centers for Disease Control and Prevention, Atlanta, Georgia, USA*

9

10

11 Corresponding author

12 Claire Philippiat

13 Institute for Advanced Biosciences,

14 Site Santé, Allée des Alpes,

15 38700, La Tronche, Grenoble, France.

16 Telephone: +33 4 76 54 94 51.

17 Email: claire.philippiat@inserm.fr

18

19 Running head: How to combine repeated urine samples.

20

21

22

23

24 ABSTRACT

25 Background: In biomarker-based studies, collecting repeated biospecimens per participant can
26 decrease measurement error, particularly for biomarkers displaying high within-subject
27 variability. Guidelines to combine such repeated biospecimens do not exist.

28 Aims: To compare the efficiency of several designs relying on repeated biospecimens to
29 estimate exposure over 7 days.

30 Methods: We quantified triclosan and bisphenol A (BPA) in all urine voids (N=427) collected
31 over seven days from eight individuals. We estimated the volume-weighted concentrations for
32 all urine samples collected during a week and compared these gold standards with the
33 concentrations obtained for equal-volume pools (standardized or not for urine dilution),
34 unequal-volume pools (based on sample volume or creatinine concentration), and for the mean
35 of the creatinine-standardized concentrations measured in each spot sample.

36 Results: For both chemicals, correlations with gold standards were similar for equal- and
37 unequal-volume pooling designs. Only for BPA, correlation coefficients were markedly lower
38 after standardization for specific gravity or creatinine of concentrations estimated in equal-
39 volume pools. Averaging BPA creatinine-standardized concentrations measured in each spot
40 sample led also to lower correlations with gold standards compared to those obtained for
41 unstandardized pooling designs.

42 Conclusion: For BPA and triclosan, considering individual urine sample volume or creatinine
43 concentrations when pooling is unnecessary because equal-volume pool adequately estimates
44 concentrations in gold standards. Standardization for specific gravity or creatinine of the
45 concentrations assessed in equal-volume pool as well as averaging creatinine-standardized
46 concentrations measured in each individual spot sample are not suitable for BPA. These results
47 provide a practical framework on how to combine repeated biospecimens in epidemiological
48 studies.

49 Keywords: Biomarkers; bisphenol A; measurement error; pooling designs; repeated
50 biospecimens

51

52 Sources of funding: Funding was provided by the U.S. government to CDC. Claire Philippiat
53 is supported by ANR (19-CE36-0003-01) and The French National Research Program for
54 Environmental and Occupational Health of Anses (2019/1/039).

55

56 Conflicts of interest: None

57 Acknowledgements: We thank Andreas Sjodin for designing the original study to evaluate
58 temporal variability of polycyclic aromatic hydrocarbon biomarkers, measuring specific
59 gravity, and providing the urine samples, the late Xiaoyun Ye for the bisphenol A and triclosan
60 biomarkers data, and Pam Olive for the creatinine measures. We thank Paulina Jedynak for her
61 useful comments on the manuscript and Matthieu Rolland for the design of figures.

62 Disclaimer: The findings and conclusions in this manuscript are those of the authors and do not
63 necessarily represent the official position of the Centers for Disease Control and Prevention.
64 Use of trade names is for identification only and does not imply endorsement by the CDC, the
65 Public Health Service, or the U.S. Department of Health and Human Services.

66

67

68 Abbreviations:

69 BPA: bisphenol A

70 CBP: creatinine-based pool

71 EVP: equal-volume pool

72 LOD: limit of detection

73 VBP: volume-based pool

74 SG: specific gravity

75

76 BACKGROUND

77 Most of the epidemiological studies on the health effects of short biological elimination half-
78 lives chemicals (e.g., bisphenol A (BPA), phthalates, triclosan, pyrethroids) have relied on a
79 relatively small number of urine samples per participant to assess exposure. Because these
80 biomarkers present high intra-individual variability in urine concentrations (reviewed by Casas
81 et al. 2018), snapshot assessments imperfectly reflect the average exposure over a day or longer
82 time periods, leading to classical measurement error and biased effect estimates (Brunekreef et
83 al. 1987; Perrier et al. 2016). Several recent cohorts aiming to reduce measurement error and
84 associated bias in effect estimates (LaKind et al. 2019; Perrier et al. 2016), collected multiple
85 biospecimens per participant during target exposure windows (Lyon-Caen et al. 2019; Shin et
86 al. 2019; Warembourg et al. 2019).

87
88 For chemicals with high intra-individual variability (e.g., intraclass correlation coefficient
89 (ICC) ≤ 0.2), up to 30 samples per subject may be required to adequately estimate a long-term
90 exposure, such as during pregnancy (Perrier et al. 2016; Vernet et al. 2019). To limit assay
91 costs, one can, for each participant, pool an equal volume of each spot sample collected over
92 the period of interest and analyze the pool instead of individual samples (Schisterman and
93 Vexler 2008; Vernet et al. 2019; Weinberg and Umbach 1999). Assuming no pooling error,
94 biomarker concentrations obtained from equal-volume pooling are equivalent to the average of
95 concentrations in each spot sample. However, such assumption is not valid when applying a
96 standardization to account for urine dilution; which often involves dividing the biomarker
97 concentration by the creatinine concentration. Averaging the creatinine-standardized
98 concentrations measured in each spot sample will lead to different concentrations than those
99 obtained by dividing the biomarker by the creatinine concentrations assessed in the pool of
100 these samples (Rosen Vollmar et al. 2020). In addition, the equal-volume pooling design does

101 not account for variations in urine dilution across spot samples and may over-represent spot
102 collections with the smallest volumes. How this over-representation may affect biomarker
103 concentrations compared to other pooling designs relying on unequal volumes of biospecimens
104 (Weinberg et al. 2019) has not been studied yet.

105

106 OBJECTIVES

107 Our aim was to compare the efficiency of several designs relying either on pooling spot urine
108 biospecimens or on averaging biomarker concentrations measured in multiple spot samples to
109 estimate concentrations in a volume-weighted concentration for all urine samples collected
110 during a week (or 24 hour) deemed to be the gold standard. We considered two chemicals with
111 contrasted intra-individual variability and correlation patterns with creatinine and specific
112 gravity: triclosan and BPA.

113

114 METHODS

115 *Study population*

116 Details regarding the study population have been described before (Li et al. 2010; Preau et al.
117 2010; Ye et al. 2011). Briefly, in October-November 2005, eight adults (four males, four
118 females) between 26 and 58 years of age, healthy, nonsmokers, and living in the metropolitan
119 Atlanta area in Georgia (USA) were recruited to participate in a study. The study, designed to
120 examine the temporal variability in urinary concentrations of polycyclic aromatic hydrocarbon
121 biomarkers, was approved by the Centers for Disease Control and Prevention (CDC)
122 Institutional Review Board. All participants signed an informed consent.

123

124 *Urine collection*

125 The study research team provided non-vinyl, non-polycarbonate plastic urine collection cups.
126 Participants were asked to collect all urine voids produced over a week and to record for each
127 void the volume and collection time. Study participants collected a total of 427 urine specimens
128 and missed 23 samples (Ye et al. 2011). Among the 427 samples collected, 10 had no volume
129 recorded and were excluded from this study. Participants decanted approximately 50 mL of
130 urine to a prelabeled, sterile, polypropylene/polyethylene urine collection cup and stored it in
131 an ice cooler containing frozen ice packs until collection by the study staff (daily or after the
132 weekend). Urine was then aliquoted into polypropylene cryovials and frozen at -70°C until
133 analysis. Using the time elapsed between two urine voids (t) and the volume of the second void
134 (V_i), we computed an urinary flow rate (UFR_i) (Middleton et al. 2016):

$$135 \quad UFR_i = V_i/t$$

136 *Quantification of BPA, triclosan, specific gravity (SG) and creatinine*

137 Urine samples were analyzed at the CDC using a method based on online solid phase extraction
138 (SPE) coupled to high performance liquid chromatography-atmospheric pressure chemical
139 ionization-isotope dilution tandem mass spectrometry (HPLC-APCI-MS/MS)(Ye et al. 2005).
140 Briefly, 100 μ L of urine spiked with the appropriate reagents and standards was incubated to
141 hydrolyze the biomarkers urinary conjugates. The procedure for extracting the deconjugated
142 biomarkers from the urine involved concurrent online SPE-HPLC operation with peak focusing
143 followed by APCI-MS/MS. In addition to study samples, each analytical run included high-
144 and low-concentration quality control materials (QCs) and reagent blanks to assure accuracy
145 and reliability of the data. The concentrations of the QCs were evaluated using standard
146 statistical probability rules. The limits of detection (LODs) were 2.3 μ g/L (triclosan) and 0.4
147 μ g/L (BPA). For analysis, concentrations below the LOD were replaced by a value equal to
148 $LOD/\sqrt{2}$ (Hornung and Reed 1990). Urinary specific gravity and creatinine were measured at

149 the CDC using a handheld digital refractometer and a Roche Hitachi 912 Chemistry Analyzer
150 (Hitachi, Pleasanton, CA), respectively.

151

152 *Temporal variability*

153 We used linear mixed models with a random intercept for participant to
154 compute Intraclass Correlation Coefficients (ICCs) between concentrations
155 measured in the spot urine samples.

156

157 *Construction of the exposure proxies*

158 For each participant and both BPA and triclosan, using the biomarker urinary concentrations
159 and volume of each collected void, we constructed the exposure proxies described below.

160

161 1) Pool of the whole volume of all individual spot samples collected over a week. The
162 concentration assessed in this pool was equivalent to the concentration that would have been
163 obtained in cumulative urine voids during a week. This concentration, named the volume-
164 weighted concentration for all urine samples collected during a week, was considered as the
165 gold standard:

$$166 \quad \text{ConcA}_{\text{Gold_standard}} = \frac{\sum_{i=1}^n \text{ConcA}_i \times \text{Volume}_i}{\sum_{i=1}^n \text{Volume}_i}$$

167 where ConcA_i was the concentration of biomarker A in the urine sample i , and volume_i was
168 the sample volume.

169 2) Equal-volume pool (EVP). We simulated the biomarker concentration that would have been
170 obtained after pooling the exact same volume of each individual spot sample:

171
$$ConcA_{EVP} = \frac{\sum_{i=1}^n ConcA_i}{N}$$

172 where N was the number of spot samples collected by the participant over a week.

173 3) Unequal-volume pool (EVP)

174 3.1) Volume-based pool (VBP). In this pool, the volume of each spot sample was equal to the
 175 ratio of its volume to the entire volume of urine collected over a week:

176
$$ConcA_{VBP} = \sum_{i=1}^n \frac{Volume_i}{\sum_{i=1}^n Volume_i} \times ConcA_i$$

177 When all the samples collected over the period of interest are included in the pool, volume-
 178 based pool is equivalent to the gold standard.

179 3.2) Creatinine-based pool (CBP, (Weinberg et al. 2019)). In this case, the volume of each spot
 180 included in the pool ($Volume_{pooled_i}$) depends on its creatinine concentration. Samples with
 181 higher creatinine concentration contribute less volume, and those with lower creatinine
 182 concentration contribute more volume:

183
$$Volume_{pooled_i} = \frac{Volume_{ref} \times Conc_{creat_{ref}}}{Conc_{creat_i}}$$

184 Where $Volume_{ref}$ and $Conc_{creat_{ref}}$ are the volume and creatinine concentration, respectively,
 185 of a selected reference spot sample whose creatinine concentration, for each participant, was
 186 near the creatinine median concentration of the spot samples collected over a week (Weinberg
 187 et al. 2019). $Conc_{creat_i}$ was the creatinine concentration in the considered spot sample. We
 188 then computed the biomarker concentration in the creatinine-based pool as follows:

189
$$ConcA_{CBP} = \frac{\sum_{i=1}^n ConcA_i \times Volume_{pooled_i}}{\sum_{i=1}^n Volume_{pooled_i}}$$

190

191 4) Equal-volume pool standardized for SG or creatinine. Because SG and creatinine
192 standardization are commonly used to account for urine dilution varying across samples, we
193 derived two additional exposure proxies from the concentrations estimated in the equal-volume
194 pools. Creatinine standardization was done by dividing the EVP biomarker concentration by
195 the EVP creatinine concentration while for SG standardization we used the following formula
196 (Philippat et al. 2013):

$$197 \quad \text{ConcA}_{EVP-SG} = \text{ConcA}_{EVP} \times [(SG_{mean_{EVP}} - 1) / (SG_{EVP} - 1)]$$

198 where $SG_{mean_{EVP}}$ was the SG arithmetic mean of the equal-volume pools in the study
199 population and SG_{EVP} equaled the SG in the considered equal-volume pool.

200 5) For each biomarker we also computed an average of the creatinine-standardized
201 concentrations measured in each spot sample. This approach did not involve pooling but has
202 been used in previous epidemiological studies that relied on biomarker concentrations obtained
203 from multiple urine samples per person (Braun et al. 2011).

204 Collecting all urine voids produced over a week can be a considerable burden for participants
205 and might limit participation rate or lead to selection bias in epidemiological studies. For this
206 reason, we also considered a scenario in which we constructed the above exposure proxies using
207 a limited number (2, 5 or 10) of voids randomly selected among the N voids collected over a
208 week for each participant.

209 *Comparison of biomarker concentrations in the exposure proxies and gold standard*

210 We relied on Spearman correlation coefficients to compare the biomarker concentrations
211 estimated for the exposure proxies (equal-volume pool standardized or not for creatinine and
212 SG, volume-based pool, creatinine-based pool and mean of the creatinine-standardized
213 concentrations measured in each spot) with the volume-weighted concentrations for all urine

214 samples collected during a week, considered as the gold standard. For the scenarios relying on
215 a limited number of randomly selected urine samples, Spearman correlation coefficients and
216 their confidence intervals were estimated using 1000 bootstraps.

217 *Sensitivity analysis*

218 In sensitivity analysis, we explored a shorter time window, specifically 24-h urine collection.
219 For each 24-h period and for each participant, we constructed the exposure proxies described
220 above and computed correlation coefficients between each exposure proxy and the urine
221 concentrations estimated in the 24-h urine collection. As for the main analysis, we considered
222 a situation where only a few (2 or 3) of the collected spot samples were included in the pool.

223 Analysis were carried out using STATA/SE, version 15.1 (StataCorp, College
224 Station, TX, USA) and R version 4.0.2. The code is available in the public repository
225 of the Team of Environmental Epidemiology applied to Reproduction and Respiratory Health
226 (<https://gricad-gitlab.univ-grenoble-alpes.fr/iab-env-epi>).

228 RESULTS

229 Out of the 427 urine samples collected, we excluded 10 (2%) from the analyses because of
230 missing volume data. The number of samples with volume information available ranged
231 between 27 and 68, depending on the participant. The average volume per void and for a 24-h
232 urine collection were 417 mL (Standard Deviation (SD): 267) and 2017 mL (SD: 832),
233 respectively. We detected BPA and triclosan in 91% and 72% of the spot samples, respectively.
234 Median concentrations in the spot samples were 1.7 (25th, 75th percentiles: 0.8, 3.6) and 9.4 (<
235 LOD, 32.8) µg/L for BPA and triclosan, respectively. Biomarker urinary concentrations
236 assessed in spots are displayed for each participant in Figure 1. ICC were relatively low for

237 BPA (ICC = 0.14, 95%CI: 0.05; 0.33), creatinine (ICC = 0.21, 95%CI: 0.08; 0.44) and SG (ICC
238 = 0.21, 95%CI: 0.08; 0.44) and relatively high for triclosan (ICC = 0.77, 95%CI: 0.56; 0.90).

239

240 *Within sample correlations across markers assessed in the study*

241 UFR was negatively correlated with SG ($\rho = -0.80$), creatinine ($\rho = -0.87$), BPA ($\rho =$
242 -0.62), and triclosan ($\rho = -0.41$). Void volume was also negatively associated with these
243 biomarkers, however the absolute values of the correlation coefficients were lower ($\rho = -0.41,$
244 $-0.37, -0.28, -0.24$ for SG, creatinine, BPA, and triclosan, respectively) than those observed for
245 the UFR. Time elapsed since the last void, SG and creatinine were all positively correlated with
246 both BPA and triclosan concentrations; the absolute value of the correlation coefficient was
247 higher for triclosan than for BPA (Table 1).

248

249 *Equal-volume pool*

250 For all participants, but one for triclosan and two for BPA, equal-volume pool estimated
251 concentrations were higher than the volume-weighted concentrations for all urine samples
252 collected during a week by 1% to 42%, depending on the participant (Figure 2). For both
253 triclosan and BPA, when all the collected spot samples were included in the equal-volume
254 pools, correlations with the gold standard were high ($\rho = 0.98$ for triclosan and 0.93 for BPA).
255 Limiting the number of samples used in the equal-volume pools had little impact for triclosan:
256 correlation coefficients between equal-volume pools and the volume-weighted concentrations
257 for all urine samples collected during a week were > 0.90 , regardless of the number of samples
258 (2, 5 or 10) included in the pool (Table 2). Limiting the number of urine samples included in
259 equal-volume pools had a stronger impact for BPA: correlation with gold standard was 0.47
260 (95% confidence interval (CI): $-0.12; 0.88$), 0.63 (95%CI: $0.10; 0.95$) and 0.75 (95%CI: $0.38;$
261 0.95) when 2, 5 and 10 samples were included, respectively.

262

263 *Unequal-volume pool*

264 *Volume-based pool:* Regardless of the number of samples considered in the volume-based
265 pools, correlation with the gold standard was high for triclosan ($\rho \geq 0.93$) and moderate to
266 high (ranged between 0.45 (95%CI: -0;17; 0.88) when two samples were used to 0.70 (95%CI:
267 0.29; 0.95) when 10 samples were used) for BPA. Our results suggested that considering void
268 volumes when pooling had little impact on biomarker concentration estimates. Spearman
269 correlations with the gold standard for volume-based pools were indeed similar (triclosan, ρ
270 ≥ 0.93) or slightly lower compared to those of equal-volume pools. For BPA, when 10 samples
271 were included in the pool, correlations with the gold standard were 0.70 (95%CI: 0.29; 0.95)
272 for volume-based pools and 0.75 (95%CI: 0.38; 0.95) for equal-volume pools, respectively.

273

274 *Creatinine-based pool:* Biomarker concentrations in the creatinine-based pools were overall
275 lower than the volume-weighted concentrations for all urine samples collected during a week
276 (except for two participants for triclosan and one for BPA (Figure 2)). Compared to the equal-
277 volume pool design, creatinine-based pools only slightly increased correlation coefficients with
278 gold standards for BPA, while both approaches gave similar results for triclosan (Table 2). For
279 example, for BPA, when 10 samples were included in the pool, correlation coefficients with
280 volume-weighted concentrations for all urine samples collected during a week were 0.79
281 (95%CI: 0.48; 0.98) for creatinine-based pools and 0.75 (95%CI: 0.38; 0.95) for equal-volume
282 pools.

283

284 *Standardization of equal-volume pools for urine dilution using creatinine or SG:*

285 Creatinine and SG standardization had little impact for triclosan. Regardless of the number of
286 spot samples considered, creatinine- and SG-standardized equal-volume pools led to similar

287 (all samples used) or slightly lower (limited number of samples included in the pool) Spearman
288 correlations with the gold standard than those observed for the unstandardized equal-volume
289 pools (all $\rho \geq 0.90$, Table 2). The impact of creatinine and SG standardization was more
290 pronounced for BPA (Table 2). Regardless of the number of samples considered, correlation
291 coefficients with gold standard were markedly weaker with than without standardization (Table
292 2). For example, when using 10 samples, Spearman correlation coefficients with the volume-
293 weighted concentrations for all urine samples collected during a week were 0.22 (95%CI: -
294 0.31; 0.71) and 0.37 (95%CI: -0.17; 0.81) for the creatinine- and SG-standardized equal-volume
295 pools, respectively, compared to 0.75 (95%CI: 0.98; 0.95) for the non-standardized equal-
296 volume pools. Overall, for both biomarkers, correlation coefficients with gold standard were
297 slightly higher for the SG-standardized than for the creatinine-standardized equal-volume
298 pools.

299

300 *Assessing concentrations in each spot and using the average instead of pooling*

301 Compared to all the pooling designs evaluated, assessing biomarker and creatinine
302 concentrations in each spot sample and using the average of the creatinine-standardized
303 concentrations led to weaker correlations with the gold standard for BPA (all $\rho \leq 0.14$ (Table
304 2)). For triclosan, correlation coefficients obtained with this design were high (≥ 0.91 ,
305 regardless of the number of samples used) and similar to those obtained with the evaluated
306 pooling designs.

307

308 *Sensitivity analysis*

309 Overall, relying on the 24-h urine collection instead of weekly urine collection as the gold
310 standard, led to similar conclusions. While concentration distributions overlapped, when all the
311 collected spot samples were included in the pools, triclosan and BPA medians tended to be

312 higher in equal-volume pools and lower in creatinine-based pools compared to the 24-h urine
313 collection (Figure 3). Pooling unequal spot volumes (based on their original creatinine
314 concentration or volume) gave equivalent correlation coefficients with the 24-urine collection
315 to equal-volume pools (Table 3). For BPA, standardization of equal-volume pools for SG or
316 creatinine or using the average of the creatinine-standardized concentrations in each spot
317 markedly decreased correlation coefficients with the 24-h urine collection compared to the
318 other tested designs.

319

320 DISCUSSION

321 In this study, we used correlation coefficients to compare urinary concentrations estimated from
322 several designs (based on pooling or averaging of the creatinine-standardized concentration
323 assessed in spot samples) with the volume-weighted concentrations for all urine samples
324 collected during a week or a day. For both BPA and triclosan, we observed similar correlations
325 with these gold standards for equal-volume, volume-based and creatinine-based pools,
326 suggesting that accounting for sample volumes or creatinine concentrations when pooling
327 might not be necessary. In addition, for BPA, a chemical with high within-subject variability
328 (ICC = 0.14) and high correlations with creatinine and SG ($\rho \geq 0.57$) equal-volume pools
329 standardized for SG and creatinine as well as averaging of the creatinine-standardized
330 concentrations assessed in each spot should be avoided because for these designs correlations
331 with the gold standard were considerably lower than those obtained for unstandardized pooling
332 designs. Such standardization had little impact for triclosan, a chemical characterized by a
333 moderate intra-individual variability ((ICC = 0.77) and moderate correlation with SG and
334 creatinine ($\rho \leq 0.37$).

335

336 *Pooling of the equal or unequal spot sample volumes*

337 For both BPA and triclosan, regardless of the number of samples included in the pool,
338 correlations with the gold standard were comparable for equal-volume, volume-based and
339 creatinine-based pools, suggesting that these three pooling designs are equivalent for estimating
340 the volume-weighted concentrations for all urine samples collected during a week or a day.
341 Compared to equal-volume pools, unequal-volume pools are more prone to technical errors
342 because the volume of each spot included in the pool varies according to its original volume or
343 creatinine concentrations. Another limitation of the creatinine-based pooling design is its cost
344 as the quantification of creatinine concentration in each spot is needed before pooling. For these
345 practical reasons, equal-volume pools might be preferred over creatinine- and volume-based
346 pools in the framework of epidemiological studies relying on biomarker assessments.
347 Noteworthy, while equal-volume, volume-based and creatinine-based pools overall preserved
348 the ranking of the individuals compared to the volume-weighted concentrations for all urine
349 samples collected during a week or a day (i.e. correlation coefficients > 0.90 when all the spot
350 samples were included in the pool), the absolute biomarker concentrations varied. When all the
351 spot samples collected were included in the pool, equal-volume pools tended to overestimate
352 the urinary concentrations while creatinine-based pools underestimated them compared to the
353 gold standards. This important fact should be considered when comparing urinary
354 concentrations across studies which have relied on different pooling designs.

355

356 *Number of samples included in the pool*

357 As expected, regardless of the design used, when a limited number of urine samples was
358 included in the pools, correlation coefficients with the gold standard were higher for triclosan
359 than for BPA, a compound showing rather high intra-individual variability. For triclosan, a pool
360 of as few as two samples adequately represented the volume-weighted concentration for all
361 urine samples collected during a day or a week (correlation coefficients > 0.92) in our study

362 population. For BPA, such high correlation was never achieved suggesting that more than 10
363 urine samples were needed to correctly estimate exposure over seven days, and more than three
364 urine samples to estimate exposure over a day. This finding is in line with a previous simulation
365 study suggesting that for a chemical with high intra-individual variability such as BPA, about
366 35 individual urine samples would be required to reduce bias in effect estimates to $< 10\%$ when
367 studying associations with a continuous outcome (Perrier et al. 2016). Casas et al. estimated
368 that four pools of 20 spot samples each would be needed to properly estimate (defined as an
369 $ICC \geq 0.80$) women exposure to BPA during a nine-month period (Casas et al. 2018).

370

371 *Standardization of equal-volume pools for creatinine or SG*

372 While standardization of the equal-volume pools for creatinine or SG had little impact for
373 triclosan in our study, for BPA correlations with the gold standard drastically dropped. These
374 results suggested that standardization in equal-volume pools was inappropriate for BPA. Of
375 note, this result was in agreement with a study comparing standardized equal-volume pools and
376 creatinine-based pools in the framework of a case control study design (Weinberg et al. 2019).
377 Using simulated data, Weinberg et al. reported a lower confidence interval coverage (i.e.,
378 proportion of simulated datasets where the confidence interval of the predicted effect estimate
379 included the true effect) for the standardized equal-volume pools than for creatinine-based
380 pools.

381

382 *Averaging of the standardized biomarker concentrations quantified in each spot sample*

383 Quantifying biomarker and creatinine concentrations in several spot samples per individual
384 allows to assess intra-individual variability (Vernet et al. 2018; Ye et al. 2011) and can be used
385 in models such as regression calibration and SIMEX to correct an exposure-health outcome
386 association for measurement error (Carroll et al. 1995). Despite the fact that such models limit

387 bias in the effect estimates, their use is still rare in biomarkers-based studies (Jackson-Browne
388 et al. 2018) and an average of the biomarker concentrations measured in each spot has been
389 sometimes used as a proxy of exposure (Braun et al. 2011; Philippat et al. 2018; Shin et al.
390 2018). When no standardization for creatinine is performed and assuming no error during
391 preparation of the pools or chemicals assessments, such approach is equivalent to the equal-
392 volume pool and considerably limits measurement error compared to the situation when a spot
393 sample is used (Perrier et al. 2016; Vernet et al. 2019). However, in our study, using the average
394 of the creatinine-standardized instead of the crude concentration led to poor correlation
395 coefficients with the BPA concentration estimated in the gold standard. This suggested that this
396 approach should be used with caution.

397

398 *Strengths and limitations*

399 Strengths of the study include the assessment of urine volume void and measurement of
400 exposure biomarkers, creatinine and SG in all spot urine samples collected over a week. The
401 average volume of the 24-h urine collection in our study population (2017 mL (SD: 832)) was
402 similar to that reported in a subsample of the 2013 National Health and Nutrition Examination
403 Survey participants (Terry et al. 2016) suggesting that the participants correctly recorded their
404 individual void volumes. Such detailed data are quite rare and challenging to collect. The
405 downside of this extensive data collection is the sample size (N = 8) which limits result
406 generalization to other populations. In addition, our conclusions only apply to biomarkers with
407 similar intra-individual variability and correlation with creatinine and SG as BPA and triclosan.
408 We empirically estimated the urinary concentrations that would have been observed in
409 theoretical pools from the urinary concentrations quantified in the collected spot samples.
410 Because our approach excluded processing errors potentially introduced by the pooling process
411 (Lyles et al. 2015), correlation coefficients estimated in our analyses might have been

412 overestimated. However, such errors are more likely to occur in pools relying on different urine
413 volumes compared to those relying on equal volumes. Therefore, even if such errors would
414 have taken place, they should not have affected our main findings suggesting that equal-volume
415 pools were as efficient as creatinine-based and volume-based pools to estimate the volume-
416 weighted concentration for all urine samples collected during a week or a day, and that no
417 standardization for SG and creatinine should be made on equal-volume pools.

418

419 CONCLUSION

420 Our results suggest that the equal-volume pooling design performs well in estimating the
421 volume-weighted concentration for all urine samples collected during a week or a day for two
422 biomarkers, BPA and triclosan, with stark differences in terms of intra-individual variability
423 and correlation with creatinine and SG. Furthermore, standardization for SG or creatinine is not
424 recommended for equal-volume pools, at least for BPA and perhaps other chemicals with
425 similarly relatively high intra-individual variability and high correlation with SG and creatinine.
426 Last, averaging of the creatinine-standardized biomarker concentrations measured in each spot
427 sample of an individual is not suitable for BPA. These findings will help epidemiologists to
428 optimize their use of repeated urine samples in biomarker-based studies.

- 430 Aylward LL, Hays SM, Zidek A. 2017. Variation in urinary spot sample, 24 h samples, and
431 longer-term average urinary concentrations of short-lived environmental chemicals:
432 implications for exposure assessment and reverse dosimetry. *J Expo Sci Environ*
433 *Epidemiol* 27:582–590; doi:10.1038/jes.2016.54.
- 434 Braun JM, Kalkbrenner AE, Calafat AM, Yolton K, Ye XY, Dietrich KN, et al. 2011. Impact
435 of Early-Life Bisphenol A Exposure on Behavior and Executive Function in Children.
436 *Pediatrics* 128:873–882; doi:10.1542/peds.2011-1335.
- 437 Brunekreef B, Noy D, Clausing P. 1987. Variability of exposure measurements in
438 environmental epidemiology. *American journal of epidemiology* 125: 892–8.
- 439 Carroll RJ, Ruppert D, Stefanski LA. 1995. *Measurement Error in Nonlinear Models, A*
440 *Modern Perspective, Second Edition*. Chapman & Hall.:London.
- 441 Casas M, Basagaña X, Sakhi AK, Haug LS, Philippat C, Granum B, et al. 2018. Variability of
442 urinary concentrations of non-persistent chemicals in pregnant women and school-aged
443 children. *Environ Int* 121:561–573; doi:10.1016/j.envint.2018.09.046.
- 444 Hornung RW, Reed LD. 1990. Estimation of Average Concentration in the Presence of
445 Nondetectable Values. *Applied Occupational and Environmental Hygiene* 5:46–51;
446 doi:10.1080/1047322X.1990.10389587.
- 447 Jackson-Browne MS, Papandonatos GD, Chen A, Calafat AM, Yolton K, Lanphear BP, et al.
448 2018. Identifying Vulnerable Periods of Neurotoxicity to Triclosan Exposure in Children.
449 *Environmental health perspectives* 126:057001; doi:10.1289/ehp2777.
- 450 Koch HM, Aylward LL, Hays SM, Smolders R, Moos R, Cocker J, et al. 2014. Inter- and intra-
451 individual variation in urinary biomarker concentrations over a 6-day samPling Period.
452 Part 2: Personal care product ingredients. *Toxicology letters* 231:261–9;
453 doi:10.1016/j.toxlet.2014.06.023.
- 454 LaKind JS, Idri F, Naiman DQ, Verner M-A. 2019. Biomonitoring and Nonpersistent
455 Chemicals-Understanding and Addressing Variability and Exposure Misclassification.
456 *Curr Environ Health Rep* 6:16–21; doi:10.1007/s40572-019-0227-2.
- 457 Lassen TH, Frederiksen H, Jensen TK, Petersen JH, Main KM, Skakkebæk NE, et al. 2013.
458 Temporal variability in urinary excretion of bisphenol A and seven other phenols in spot,
459 morning, and 24-h urine samples. *Environ Res* 126:164–170;
460 doi:10.1016/j.envres.2013.07.001.
- 461 Li Z, Romanoff LC, Lewin MD, Porter EN, Trinidad DA, Needham LL, et al. 2010. Variability
462 of urinary concentrations of polycyclic aromatic hydrocarbon metabolite in general
463 population and comparison of spot, first-morning, and 24-h void sampling. *J Expo Sci*
464 *Environ Epidemiol* 20:526–535; doi:10.1038/jes.2009.41.
- 465 Lyles RH, Van Domelen D, Mitchell EM, Schisterman EF. 2015. A Discriminant Function
466 Approach to Adjust for Processing and Measurement Error When a Biomarker is Assayed
467 in Pooled Samples. *Int J Environ Res Public Health* 12:14723–14740;
468 doi:10.3390/ijerph121114723.
- 469 Lyon-Caen S, Siroux V, Lepeule J, Lorimier P, Hainaut P, Mossuz P, et al. 2019. Deciphering
470 the Impact of Early-Life Exposures to Highly Variable Environmental Factors on Foetal
471 and Child Health: Design of SEPAGES Couple-Child Cohort. *Int J Environ Res Public*
472 *Health* 16; doi:10.3390/ijerph16203888.
- 473 Meeker JD, Cantonwine DE, Rivera-Gonzalez LO, Ferguson KK, Mukherjee B, Calafat AM,
474 et al. 2013. Distribution, Variability, and Predictors of Urinary Concentrations of Phenols
475 and Parabens among Pregnant Women in Puerto Rico. *Environmental science &*
476 *technology* 47:3439–47; doi:10.1021/es400510g.
- 477 Middleton DRS, Watts MJ, Lark RM, Milne CJ, Polya DA. 2016. Assessing urinary flow rate,

478 creatinine, osmolality and other hydration adjustment methods for urinary biomonitoring
479 using NHANES arsenic, iodine, lead and cadmium data. *Environ Health* 15:68;
480 doi:10.1186/s12940-016-0152-x.

481 Perrier F, Giorgis-Allemand L, Slama R, Philippat C. 2016. Within-subject Pooling of
482 Biological Samples to Reduce Exposure Misclassification in Biomarker-based Studies.
483 *Epidemiology (Cambridge, Mass)* 27:378–88; doi:10.1097/ede.0000000000000460.

484 Philippat C, Barkoski J, Tancredi DJ, Elms B, Barr DB, Ozonoff S, et al. 2018. Prenatal
485 exposure to organophosphate pesticides and risk of autism spectrum disorders and other
486 non-typical development at 3 years in a high-risk cohort. *International journal of hygiene
487 and environmental health*; doi:10.1016/j.ijheh.2018.02.004.

488 Philippat C, Wolff MS, Calafat AM, Ye X, Bausell R, Meadows M, et al. 2013. Prenatal
489 Exposure to Environmental Phenols: Concentrations in Amniotic Fluid and Variability in
490 Urinary Concentrations during Pregnancy. *Environmental health perspectives* 121:1225–
491 1231; doi:10.1289/ehp.1206335.

492 Pollack AZ, Perkins NJ, Mumford SL, Ye A, Schisterman EF. 2013. Correlated biomarker
493 measurement error: an important threat to inference in environmental epidemiology.
494 *American journal of epidemiology* 177:84–92; doi:10.1093/aje/kws209.

495 Preau Jr JL, Wong LY, Silva MJ, Needham LL, Calafat AM. 2010. Variability over One Week
496 in the Urinary Concentrations of Metabolites of Diethyl Phthalate and Di(2-Ethylhexyl)
497 Phthalate among 8 Adults: an Observational Study. *Environmental health perspectives*
498 118: 1748–54.

499 Rosen Vollmar AK, Johnson CH, Weinberg CR, Deziel NC, Baird DD, Wilcox AJ, et al. 2020.
500 Accounting for urinary dilution in peri-implantation samples: implications for creatinine
501 adjustment and specimen pooling. *J Expo Sci Environ Epidemiol*; doi:10.1038/s41370-
502 020-0227-1.

503 Schisterman EF, Vexler A. 2008. To pool or not to pool, from whether to when: applications
504 of pooling to biospecimens subject to a limit of detection. *Paediatr Perinat Epidemiol*
505 22:486–96; doi:10.1111/j.1365-3016.2008.00956.x.

506 Shin H-M, Bennett DH, Barkoski J, Ye X, Calafat AM, Tancredi D, et al. 2019. Variability of
507 urinary concentrations of phthalate metabolites during pregnancy in first morning voids
508 and pooled samples. *Environ Int* 122:222–230; doi:10.1016/j.envint.2018.11.012.

509 Shin H-M, Schmidt RJ, Tancredi D, Barkoski J, Ozonoff S, Bennett DH, et al. 2018. Prenatal
510 exposure to phthalates and autism spectrum disorder in the MARBLES study. *Environ
511 Health* 17:85; doi:10.1186/s12940-018-0428-4.

512 Terry AL, Cogswell ME, Wang C-Y, Chen T-C, Loria CM, Wright JD, et al. 2016. Feasibility
513 of collecting 24-h urine to monitor sodium intake in the National Health and Nutrition
514 Examination Survey. *Am J Clin Nutr* 104:480–488; doi:10.3945/ajcn.115.121954.

515 Vernet C, Philippat C, Agier L, Calafat AM, Ye X, Lyon-Caen S, et al. 2019. An Empirical
516 Validation of the Within-subject Biospecimens Pooling Approach to Minimize Exposure
517 Misclassification in Biomarker-based Studies. *Epidemiology* 30:756–767;
518 doi:10.1097/EDE.0000000000001056.

519 Vernet C, Philippat C, Calafat AM, Ye X, Lyon-Caen S, Siroux V, et al. 2018. Within-Day,
520 Between-Day, and Between-Week Variability of Urinary Concentrations of Phenol
521 Biomarkers in Pregnant Women. *Environ Health Perspect* 126:037005;
522 doi:10.1289/EHP1994.

523 Warembourg C, Basagaña X, Seminati C, de Bont J, Granum B, Lyon-Caen S, et al. 2019.
524 Exposure to phthalate metabolites, phenols and organophosphate pesticide metabolites
525 and blood pressure during pregnancy. *Int J Hyg Environ Health* 222:446–454;
526 doi:10.1016/j.ijheh.2018.12.011.

527 Weinberg CR, Shi M, O'Brien KM, Umbach DM. 2019. Adjustment for urinary creatinine or

528 serum lipids for analytes assayed in pooled specimens. *Epidemiology*;
529 doi:10.1097/EDE.0000000000001053.
530 Weinberg CR, Umbach DM. 1999. Using pooled exposure assessment to improve efficiency in
531 case-control studies. *Biometrics* 55: 718–26.
532 Ye X, Kuklennyik Z, Needham LL, Calafat AM. 2005. Automated on-line column-switching
533 HPLC-MS/MS method with peak focusing for the determination of nine environmental
534 phenols in urine. *Anal Chem* 77:5407–5413; doi:10.1021/ac050390d.
535 Ye X, Wong LY, Bishop AM, Calafat AM. 2011. Variability of urinary concentrations of
536 bisphenol A in spot samples, first morning voids, and 24-hour collections. *Environmental*
537 *health perspectives* 119:983–8; doi:10.1289/ehp.1002701.
538

539 Table 1: Within sample Spearman correlation coefficients for the different markers assessed in
 540 the current study
 541

	Void volume	Urinary flow rate	Time since last void	Specific gravity	Creatinine	BPA
Void volume	1.00					
Urinary flow rate	0.46	1.00				
Time since last void	0.31	-0.66	1.00			
Specific gravity	-0.41	-0.80	0.48	1.00		
Creatinine	-0.37	-0.87	0.59	0.90	1.00	
BPA	-0.28	-0.62	0.40	0.57	0.64	1.00
Triclosan	-0.24	-0.41	0.25	0.33	0.37	0.24

542 N = 417 urine spot samples, except for the correlations with urinary flow rate that were restricted to the
 543 400 samples with available data for time since last void. Abbreviation: BPA: bisphenol A.
 544

545

Table 2: Spearman correlation coefficients between the volume-weighted concentrations for all urine samples collected during a week (gold standard) and the concentrations estimated from different protocols using pooling of spot samples or averaging creatinine-standardized concentrations measured in each spot sample

	All voids used to construct exposure proxies		2 randomly selected voids used to construct exposure proxies				5 randomly selected voids used to construct exposure proxies				10 randomly selected voids used to construct exposure proxies			
	Triclosan	BPA	Triclosan		BPA		Triclosan		BPA		Triclosan		BPA	
	rho	rho	rho ^a	95% CI ^a	rho ^a	95% CI ^a	rho ^a	95% CI ^a	rho ^a	95% CI ^a	rho ^a	95% CI ^a	rho ^a	95% CI ^a
Equal-volume pool	0.98	0.93	0.93	[0.78; 0.99]	0.47	[-0.12; 0.88]	0.96	[0.85; 1.00]	0.63	[0.10; 0.95]	0.97	[0.92; 1.00]	0.75	[0.38; 0.95]
Volume-based pool	1.00	1.00	0.93	[0.79; 1.00]	0.45	[-0.17; 0.88]	0.96	[0.88; 1.00]	0.58	[0.07; 0.93]	0.98	[0.92; 1.00]	0.70	[0.29; 0.95]
Creatinine-based pool	0.98	0.98	0.92	[0.76; 1.00]	0.50	[-0.12; 0.90]	0.96	[0.86; 1.00]	0.67	[0.19; 0.95]	0.97	[0.92; 1.00]	0.79	[0.48; 0.98]
SG standardized equal-volume pool	0.98	0.52	0.92	[0.76; 1.00]	0.21	[-0.40; 0.74]	0.95	[0.83; 1.00]	0.29	[-0.29; 0.81]	0.96	[0.88; 1.00]	0.37	[-0.17; 0.81]
Creatinine-standardized equal-volume pool	0.98	0.45	0.90	[0.76; 1.00]	0.11	[-0.50; 0.71]	0.94	[0.81; 1.00]	0.18	[-0.36; 0.74]	0.95	[0.88; 1.00]	0.22	[-0.31; 0.71]
Average of the spot creatinine-standardized concentrations	0.98	0.12	0.91	[0.76; 1.00]	0.09	[-0.50; 0.69]	0.94	[0.83; 0.98]	0.13	[-0.43; 0.71]	0.96	[0.88; 0.98]	0.14	[-0.33; 0.69]

^a estimated using 1000 bootstraps

Abbreviation: BPA: bisphenol A, CI: confidence interval, rho: Spearman correlation coefficient, SG: specific gravity

Table 3: Spearman correlation coefficients between BPA and triclosan concentrations estimated from 24-h urine collections (gold standard) and from different protocols using pooling of spot samples or averaging of the creatinine-standardized concentrations measured in each spot sample

	All voids used to construct exposure proxies		Only 2 randomly selected voids used to construct exposure proxies				Only 3 randomly selected voids used to construct exposure proxies			
	Triclosan	BPA	Triclosan		BPA		Triclosan		BPA	
	rho	rho	rho ^a	95% CI ^a	rho ^a	95% CI ^a	rho ^a	95% CI ^a	rho ^a	95% CI ^a
Equal-volume pool	0.99	0.95	0.95	[0.94 ; 0.97]	0.72	[0.59 ; 0.82]	0.97	[0.95 ; 0.98]	0.80	[0.70 ; 0.88]
Volume-based pool	1.00	1.00	0.95	[0.93 ; 0.97]	0.72	[0.58 ; 0.83]	0.97	[0.95 ; 0.98]	0.80	[0.71 ; 0.88]
Creatinine-based pool	0.98	0.91	0.95	[0.92 ; 0.97]	0.71	[0.57 ; 0.82]	0.96	[0.94 ; 0.98]	0.79	[0.69 ; 0.86]
SG standardized equal volume pool	0.96	0.76	0.94	[0.93 ; 0.96]	0.60	[0.44 ; 0.73]	0.95	[0.93 ; 0.96]	0.66	[0.52 ; 0.77]
Creatinine-standardized equal volume pool	0.94	0.66	0.93	[0.91 ; 0.95]	0.51	[0.35 ; 0.67]	0.94	[0.92 ; 0.95]	0.57	[0.43 ; 0.69]
Average of the spot creatinine-standardized concentrations	0.95	0.61	0.93	[0.91 ; 0.95]	0.53	[0.36 ; 0.66]	0.94	[0.92 ; 0.95]	0.57	[0.43 ; 0.69]

^a estimated using 1000 bootstraps

Abbreviation: BPA: bisphenol A, rho: Spearman correlation coefficients, SG: specific gravity, CI: Confidence Interval

Figure 1: Biomarker concentrations for all participants in all spots collected over the week of urine collection

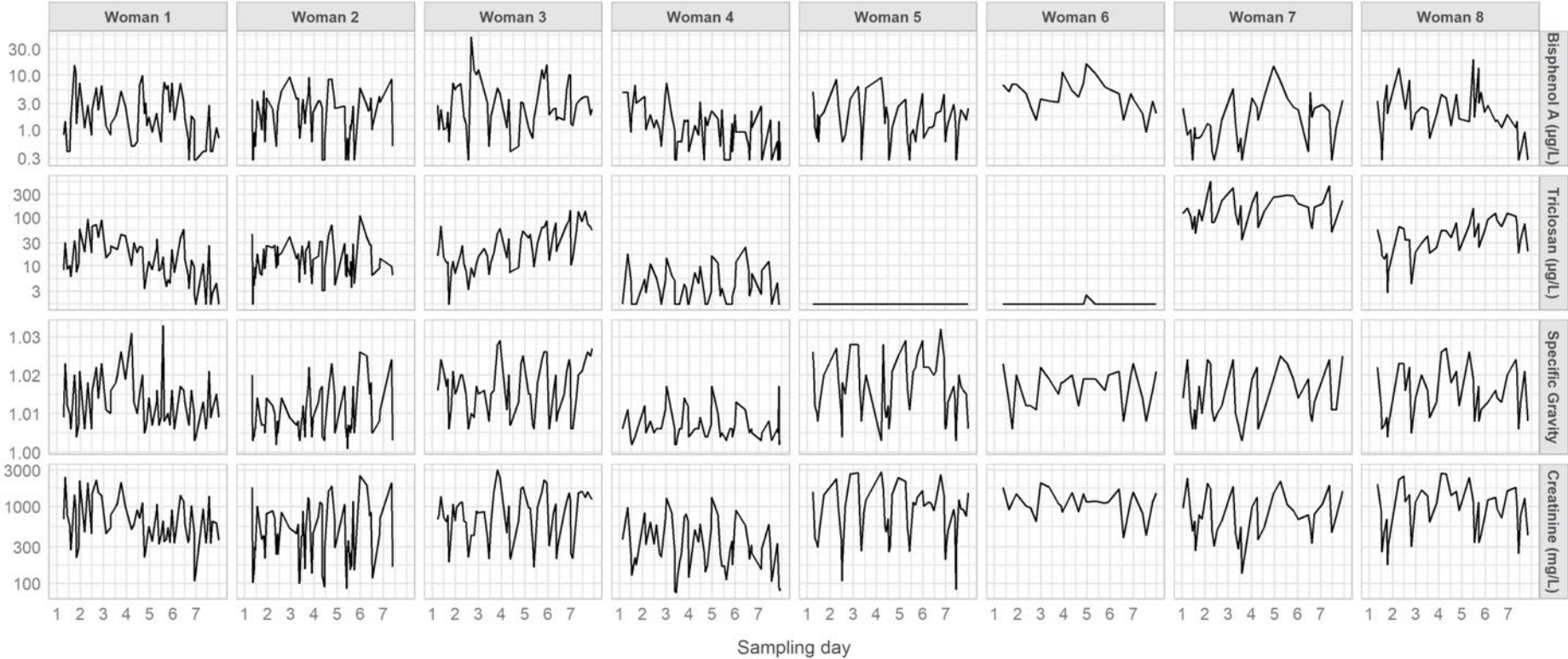


Figure 2: Estimated BPA and triclosan concentrations for each individual in gold standard (volume-weighted concentrations for all urine samples collected during a week), equal-volume (EVP) and creatinine-based (CBP) pools.

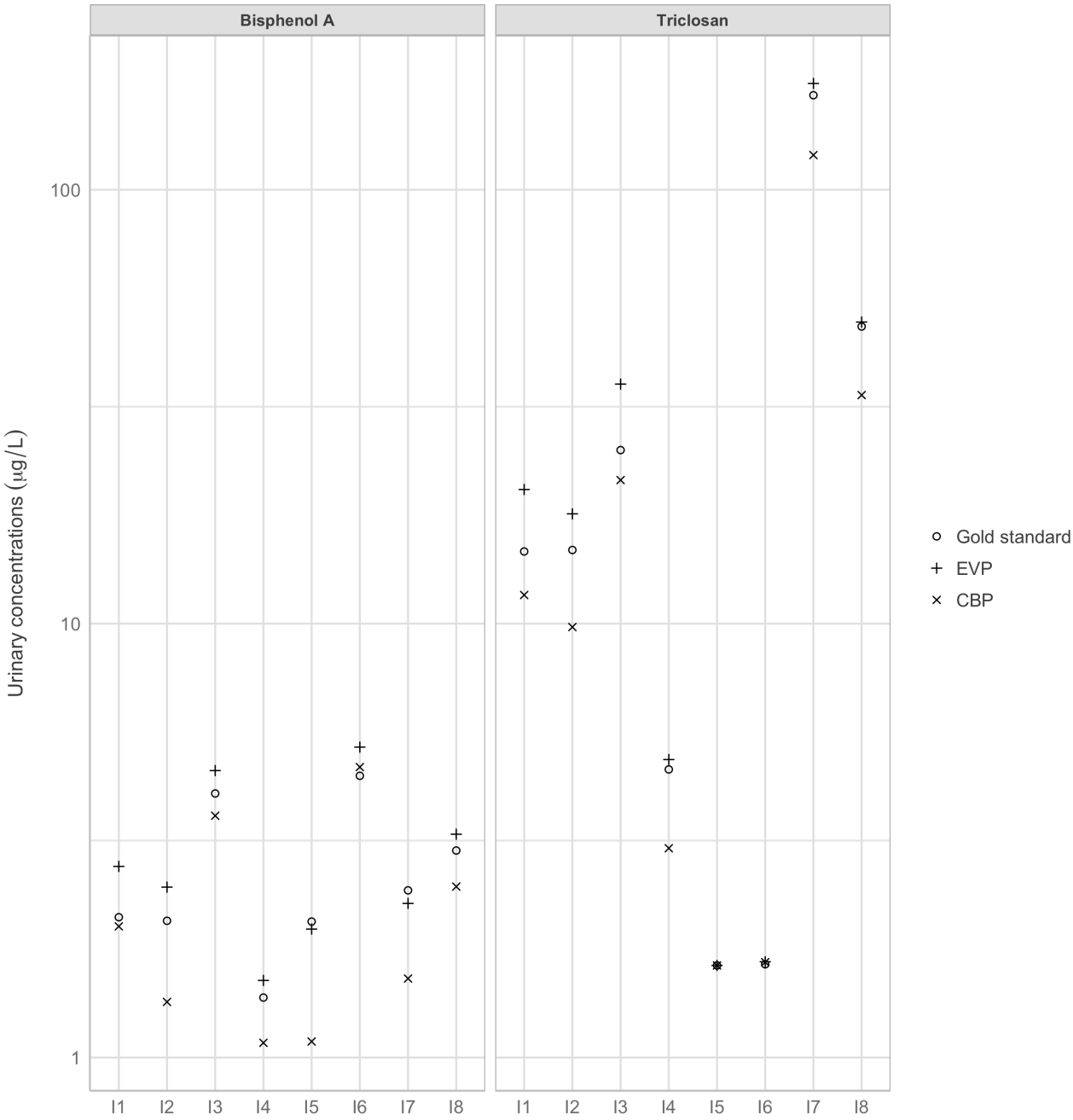
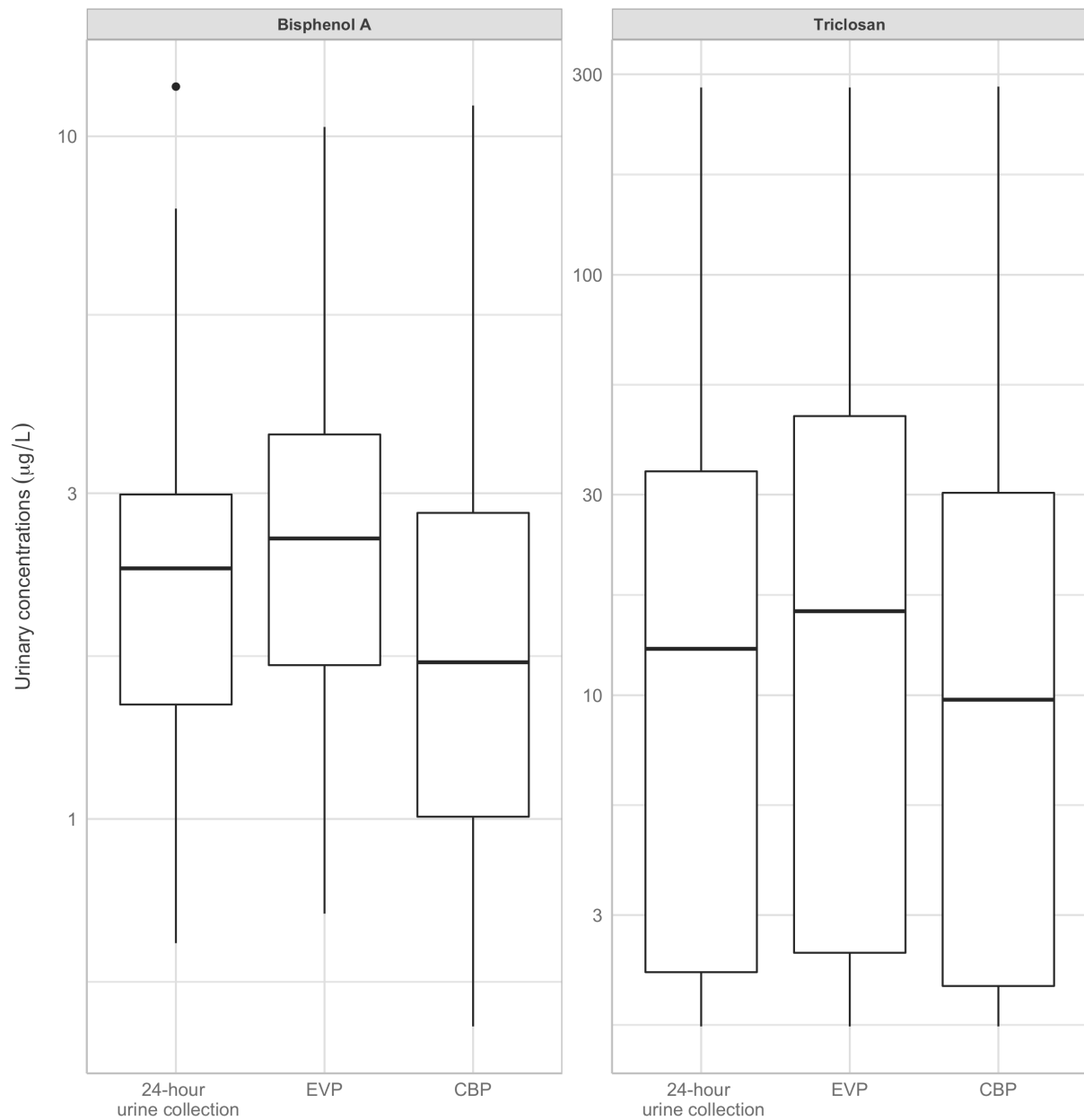


Figure 3: Distribution of BPA and triclosan concentrations estimated in 24-h urine collection (gold standard), equal-volume (EVP) and creatinine-based (CBP) pools among the eight study participants



Boxes lines represent 75th (upper line), 50th (middle line) and 25th (lower line) percentiles.